

1. What is Unsupervised Learning?

Suppose we don't give any labeled output data to an algorithm, then it will be called Unsupervised Learning. Instead, it learns patterns and relationships from the input data without explicit guidance.

2. Name some common algorithms used in Unsupervised Learning.

Some of the most frequent unsupervised learning algorithms are K-means clustering; hierarchical clustering; DBSCAN (Density-Based Spatial Clustering of Applications with Noise); Principal Component Analysis (PCA); t-Distributed Stochastic Neighbor Embedding (t-SNE).

3. Explain the difference between supervised and unsupervised learning.

In supervised learning, the algorithm is trained on labeled data, while in unsupervised learning, the algorithm learns from unlabeled data without explicit guidance.

4. What is clustering?

The procedure for bringing together similar pieces of information based on a set of attributes or characteristics is referred to as clustering.

5. Explain the K-means clustering algorithm.

Based on mean proximity, K-means divides data into K clusters during various iterations.

6. What is the elbow method in K-means clustering?

To find the optimal number of clusters for k-means by plotting variance against number of clusters and looking for an “elbow” point where rate of decrease changes sharply.

7. What are the applications of clustering?

Segmentation of customers, detection of anomalies, image segmentation, document grouping etc.

8. Explain hierarchical clustering.

Hierarchical clustering creates a tree-like structure where each branch has several leaves by merging individual points or splitting one cluster recursively.

9. What is the difference between complete-linkage and single-linkage clustering?

Complete linkage measures distance between two clusters as farthest apart points while single linkage measures distance between their nearest points.

10. What is dimensionality reduction?

Dimensionality reduction entails minimizing dimensions that do not matter in a dataset while maintaining important information within it.

11. Explain Principal Component Analysis (PCA).

PCA represents a dimensionality reduction technique that transforms data into new

coordinate systems to maximize variance along the principal components.

12. What is the curse of dimensionality?

The phrase 'curse of dimensionality' refers to the challenges and sparseness associated with dealing with high-dimensional data resulting in increased computational complexity and need for more data.

13. What is t-SNE?

t-SNE or t-distributed Stochastic Neighbor Embedding is a visualization tool used for reducing dimensions of high dimensional datasets by keeping pairwise similarities between data points.

14. Explain density-based clustering (DBSCAN).

Density-Based Spatial Clustering of Applications with Noise refers to an algorithm that groups together closely located data points having a sufficient number of neighbors, while outliers are treated as noise.

15. What is anomaly detection?

Anomaly detection involves finding rare cases, events or records deviating significantly from what is considered to be standard.

16. How does Isolation Forest work in anomaly detection?

Isolation Forest works by segregating anomalies through recursive partitioning of the given dataset, where anomalies are points that are few splits away to be isolated.

17. What is the difference between generative and discriminative models?

Generative models represent the joint probability model of features and labels, while discriminative models represent the conditional probability model for labels given observed features.

18. What is the EM algorithm?

EM stands for Expectation-Maximization which describes an iterative method used to find maximum likelihood or maximum a posteriori estimates of parameters in statistical models.

19. Explain the difference between PCA and t-SNE.

The main distinction between PCA and t-SNE lies in linearity as PCA maximizes variance linearly while t-SNE preserves local similarities in high-dimensional space non-linearly.

20. What is the difference between regression and clustering?

Regression predicts continuous output whereas clustering groups similar data points without giving any attention to their labels.

21. How do you handle missing data in unsupervised learning?

The problem can be solved using techniques like imputation or removing missing values. Imputation estimates missing values based on observed data.

22. Explain the concept of feature scaling.

Feature scaling is a process that normalizes or standardizes input features to ensure they have similar scales so that one of the features does not dominate others.

23. What is the difference between correlation and covariance?

Covariance measures the extent of joint variability between two random variables while correlation measures direction and strength of linear relationship between two variables, it has a range of -1 to 1.

24. How does the choice of distance metric impact clustering results?

The choice determines how clusters are derived. The different metrics yield different cluster assignments and which metric is more appropriate depends on the nature of the dataset.

25. What is the role of the centroid in K-means clustering?

Centroid represents the mean value for all data points in a single cluster in k-means clustering algorithm, which is utilized for iterative updating of cluster assignments.

Related posts:

1. What is Machine learning ?
2. Define machine learning and explain its importance in real-world applications.
3. What are the different types of machine learning?
4. What is a hyperparameter in machine learning ?
5. TOP INTERVIEW QUESTIONS AND ANSWERS FOR Artificial Intelligence

6. Deep Learning Top Interview Questions and Answers
7. Differences Between Machine Learning and Artificial Intelligence
8. Machine Learning works on which type of data ?
9. What is Regression in Machine learning
10. Finding Machine Learning Datasets
11. What is hypothesis function and testing
12. Explain computer vision with an appropriate example
13. Explain Reinforcement learning with an appropriate exaple
14. Reinforcement Learning Framework
15. Data augmentation
16. Normalizing Data Sets in Machine Learning
17. Machine learning models
18. Unsupervised machine learning
19. Neural Network in Machine Learning
20. Recurrent neural network
21. Support Vector Machines
22. Long short-term memory (LSTM) networks
23. Convolutional neural network
24. How to implement Convolutional neural network in Python
25. What does it mean to train a model on a dataset ?
26. Can a textual dataset be used with an openCV?
27. Name some popular machine learning libraries.
28. Introduction to Machine Learning